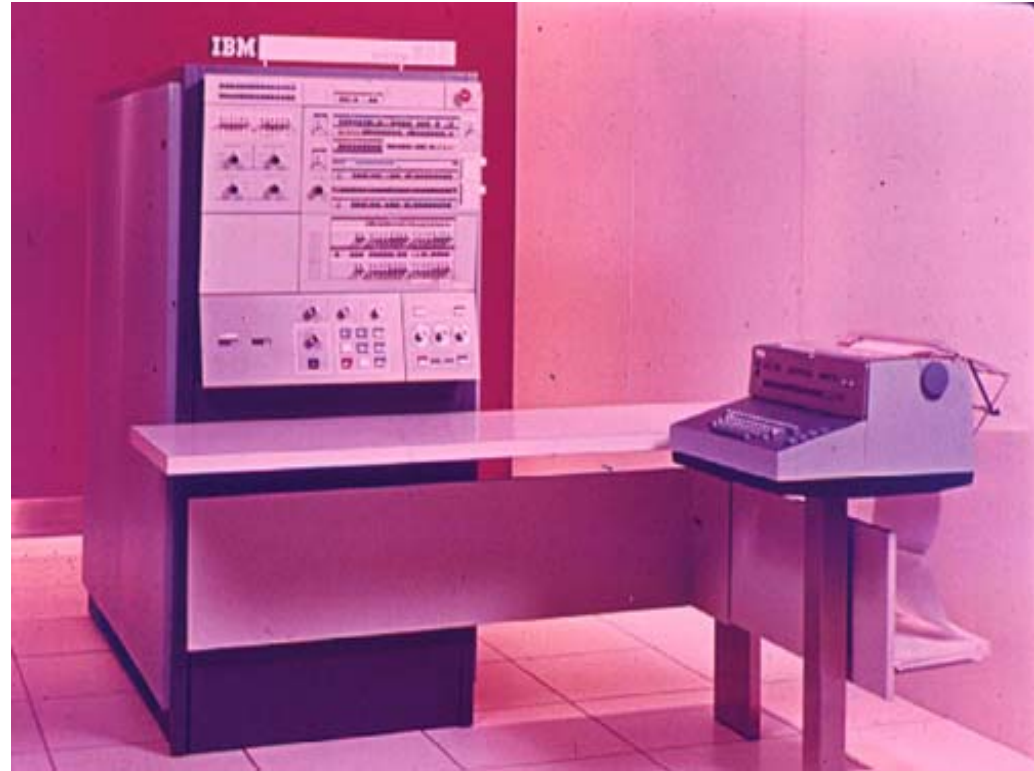
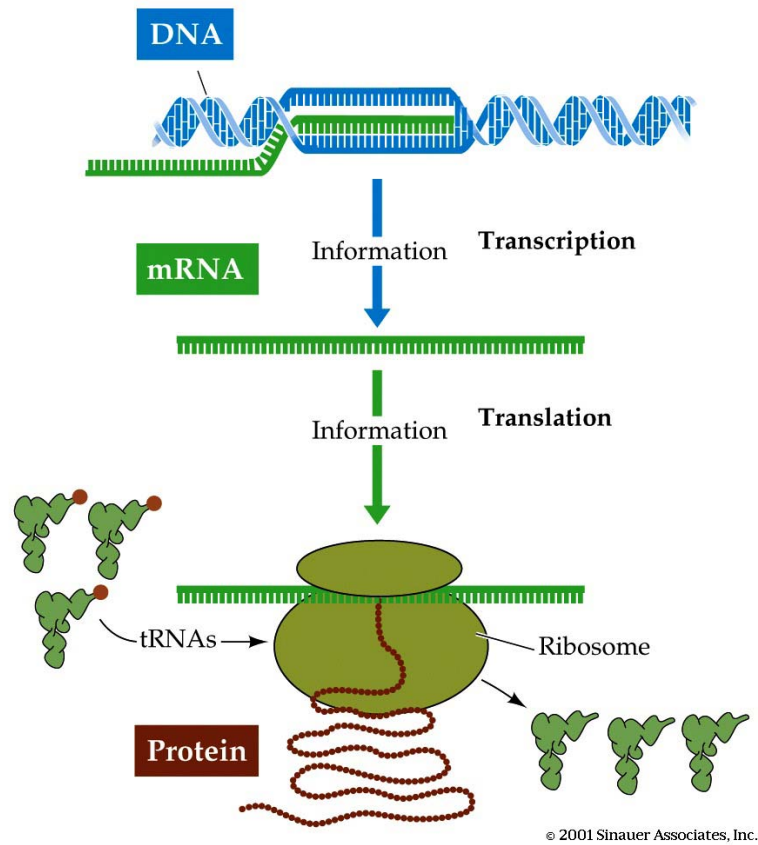


27622: Introduction to Bioinformatics, Turbo version

Henrik Nielsen, Associate professor
Center for Biological Sequence
Analysis

What is bioinformatics?



What are bioinformaticians up to, actually?

- *Manage* molecular biological data
 - Store in *databases*, organise, formalise, describe...
- *Compare* molecular biological data
- Find *patterns* in molecular biological data
 - *phylogenies*
 - *correlations* (sequence / structure / expression / function / disease)

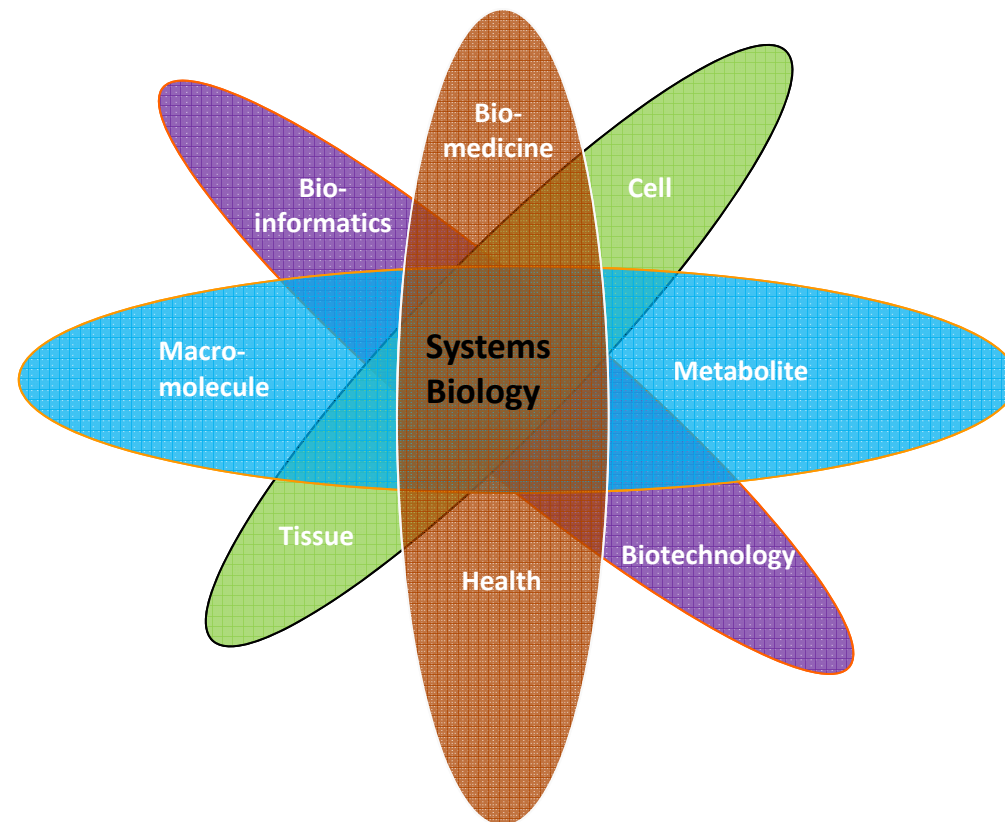
Goals:

- *characterise* biological patterns & processes
- *predict* biological properties
 - low level data \Rightarrow high level properties
(eg., sequence \Rightarrow function)

Bioinformatics: neighbour disciplines

- Computational biology
 - *Broader concept: includes computational ecology, physiology, neurology etc...*
- -omics:
 - *Genomics*
 - *Transcriptomics*
 - *Proteomics*
- Systems biology
 - *Putting it all together...*
 - *Building models, identify control & regulation*

A view of Systems Biology



- **Bio- side:**
 - Molecular biology
 - Cell biology
 - Genetics
 - Evolutionary theory
- **-informatics side:**
 - Computer science
 - Statistics
 - Theoretical physics

• DNA sequences

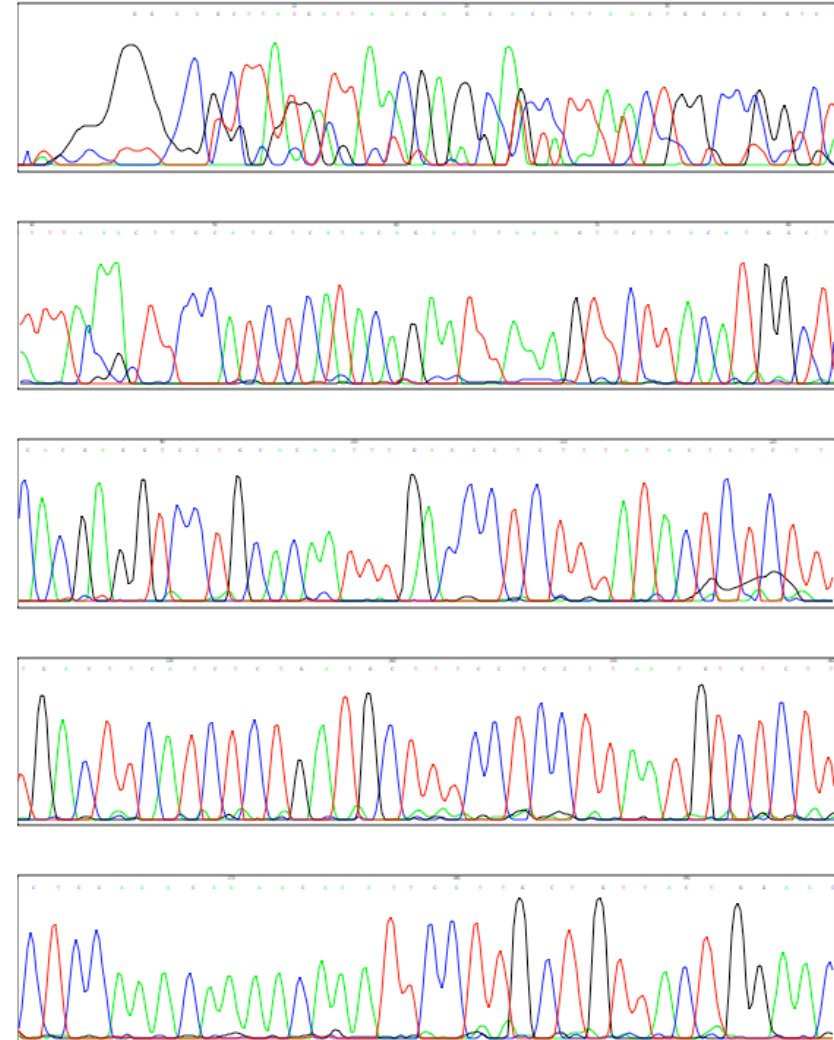
>alpha-D

```
ATGCTGACCGACTCTGACAAGAAGCTGGTCTGCAGGTGTGGGAGAAGGTGATCCGCCAC
CCAGACTGTGGAGCCGAGGCCCTGGAGAGGTGCGGGCTGAGCTTGGGGAAACCATGGGCA
AGGGGGGCGACTGGGTGGGAGCCCTACAGGGCTGCTGGGGGTTGTTCGGCTGGGGGTCAG
CACTGACCATCCCGCTCCCGCAGCTGTTACACCTACCCCAAGCAAGACCTACTTCC
CCCACTTCGACTTGCACCATGGCTCCGACCAGGTCCGCAACCAAGCAAGAGGTGTTGG
CCGCTTGGGCAACGCTGTCAAGAGCCTGGGCAACCTCAGCCAAGCCCTGTCTGACCTCA
GCGACCTGCATGCTTACAACCTGCGTGTGACCTGTCAACTTCAAGGCAGGCGGGGAC
GGGGGTCAGGGGCCGGGGAGTTGGGGGCCAGGGACCTGGTTGGGGATCCGGGGCCATGCC
GGCGGTACTGAGCCCTGTTTTCCTTGCAGCTGCTGGCGCAGTGTCTCCACGTGGTGCTG
GCCACACACCTGGGCAACGACTACACCCGGAGGCATGCTGCTTCGACAAGTTCCTG
TCGGCTGTGTGCACCCGTGCTGGCCGAGAAGTACAGATAA
```

>alpha-A

```
ATGGTGCTGTCTGCCAACGACAAGAGCAACGTGAAGGCCGTCTTCGGCAAAATCGGCGGC
CAGGCCGGTGAAGTGGGTGGTGAAGCCCTGGAGAGGTATGTGGTCATCCGTCATTACCCC
ATCTCTGTCTGTCTGTGACTCCATCCCATCTGCCCCATACTCTCCCATCCATAACTG
TCCCTGTTCTATGTGGCCCTGGCTCTGTCTCATCTGTCCCAACTGTCCCTGATTGCCTC
TGTCCTCCAGGTTGTTTCATCACCTACCCCAAGCAAGACCTACTTCCCACTTCGACC
TGTCACATGGCTCCGCTCAGATCAAGGGGCACGGCAAGAGGTGGCGGAGGCACTGGTTG
AGGCTGCCAACCATCGATGACATCGCTGGTGCCTCTCCAAAGTGAAGGACCTCCACG
CCCAAAAGCTCCGTGTGGACCCCGTCAACTTCAAAGTGAGCATCTGGGAAGGGTGACCA
GTCTGGCTCCCTTCCTGCACACACCTCTGGCTACCCCTCACCTACCCCTTGTCTCACC
ATCTCCTTTTGCCTTTCAGCTGCTGGGTCACTGCTTCTGCTGGTGGTGGCGGTCACCTT
CCCTCTCTCTGACCCCGAGGTCCATGCTTCCCTGGACAAGTTCGTGTGTGCGGTGGG
CACCGTCTTACTGCCAAGTACCGTTAA
```

File: /Users/raz/Download/byc_97801.z1.abd.txt
Run ended: 9:05AM Thu, Jun 07, 2001
(page 1)



Molecular biology data...

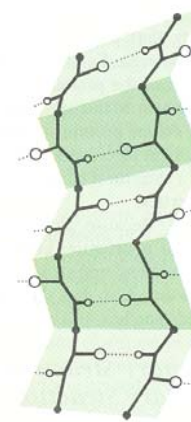
- Amino acid sequences
- Protein structure:
 - X-ray crystallography
 - NMR



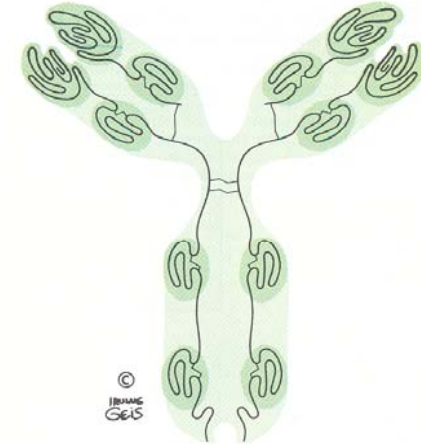
(a) Primary structure (amino acid sequence in the protein chain)



α helix



β sheet



Domains (dark color) in an antibody molecule

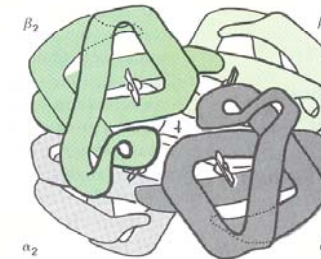
(b) Secondary structure

(c) Local folding



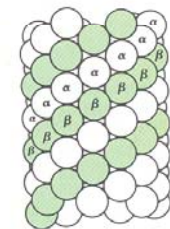
One complete protein chain (β chain of hemoglobin)

(d) Tertiary structure



The four separate chains of hemoglobin assembled into an oligomeric protein

(e) Quaternary structure



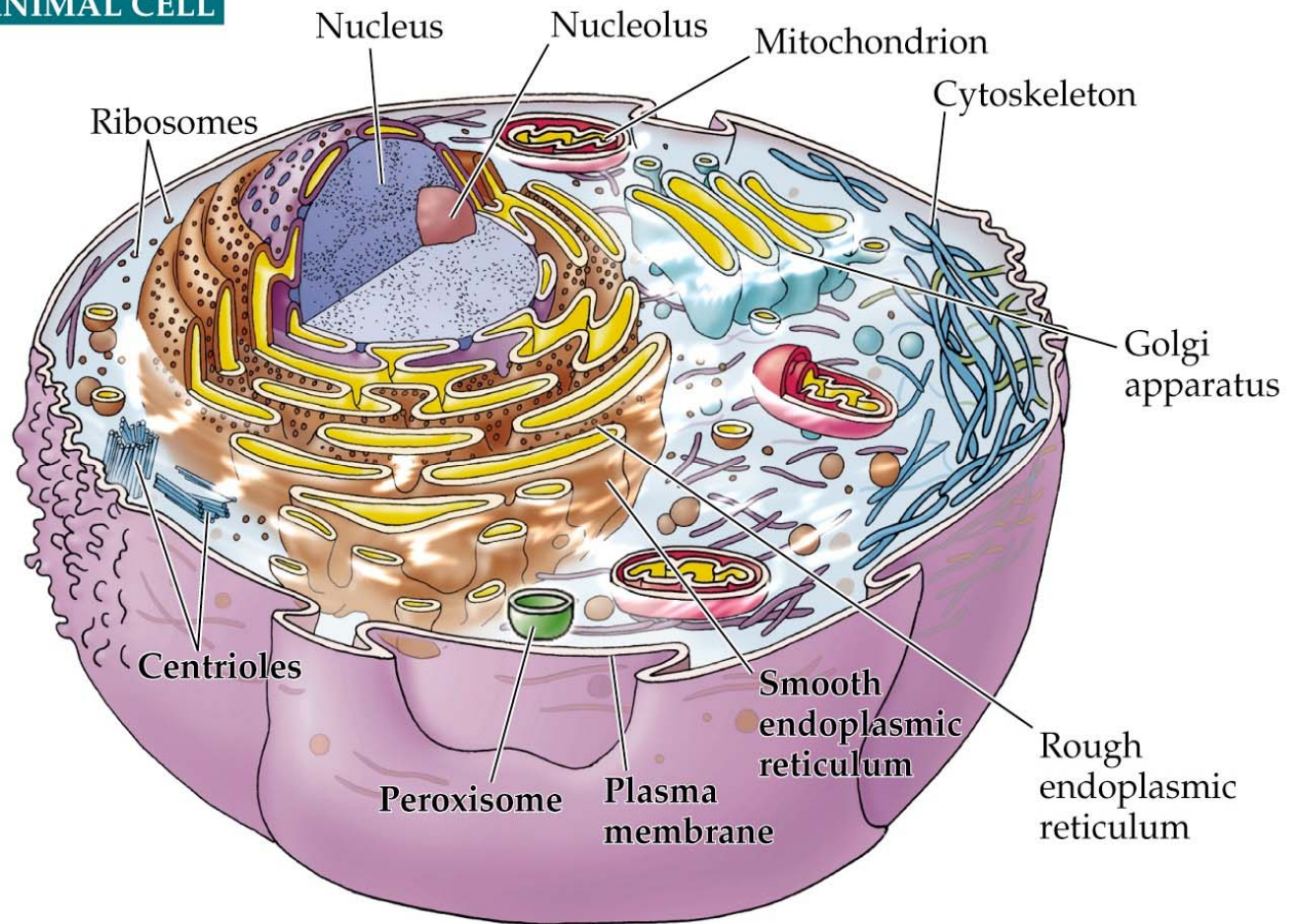
α (white) and β (color) tubulin molecules in a microtubule

(f) Quaternary structure

Cell biology & proteomics data...

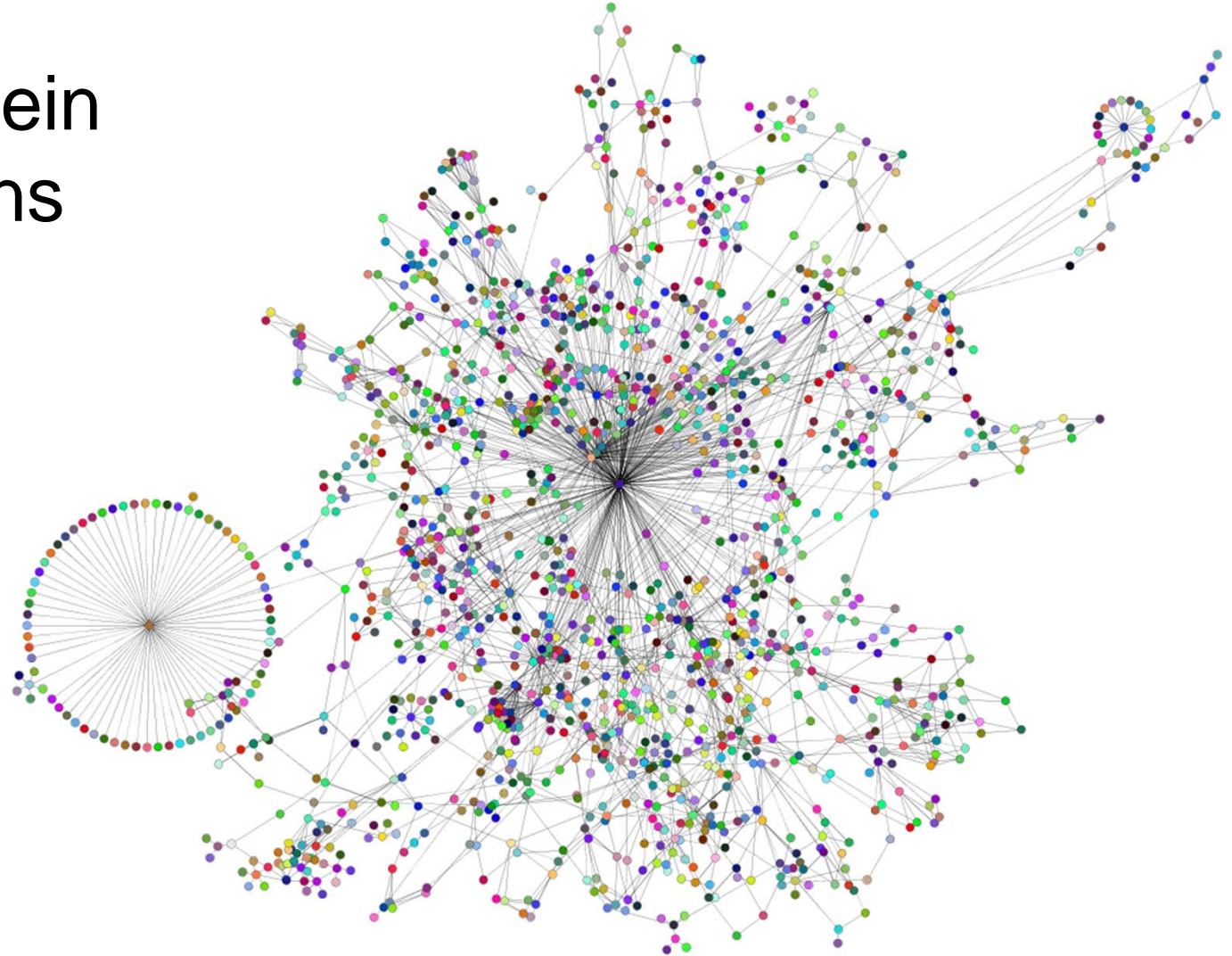
- Subcellular localization

AN ANIMAL CELL

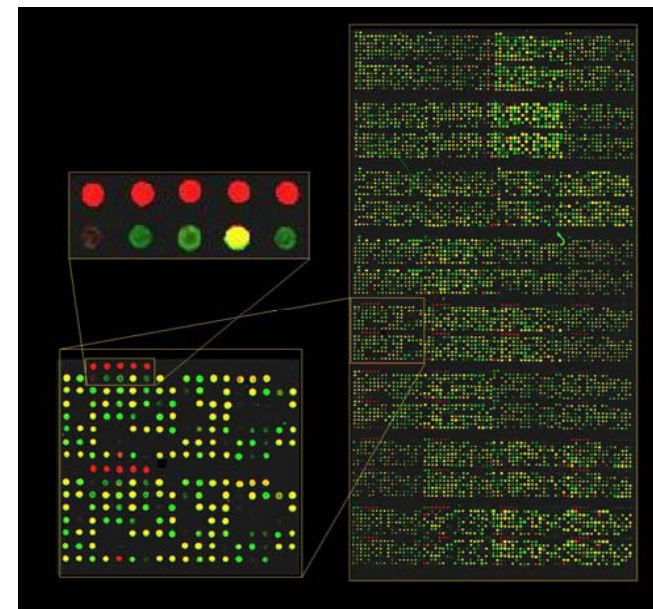
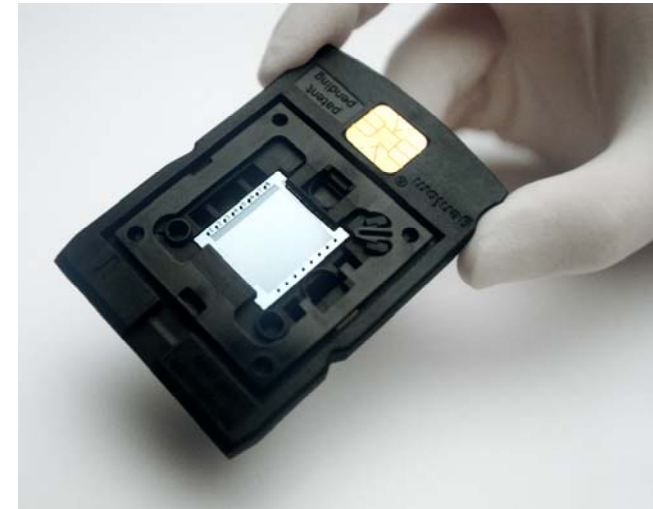
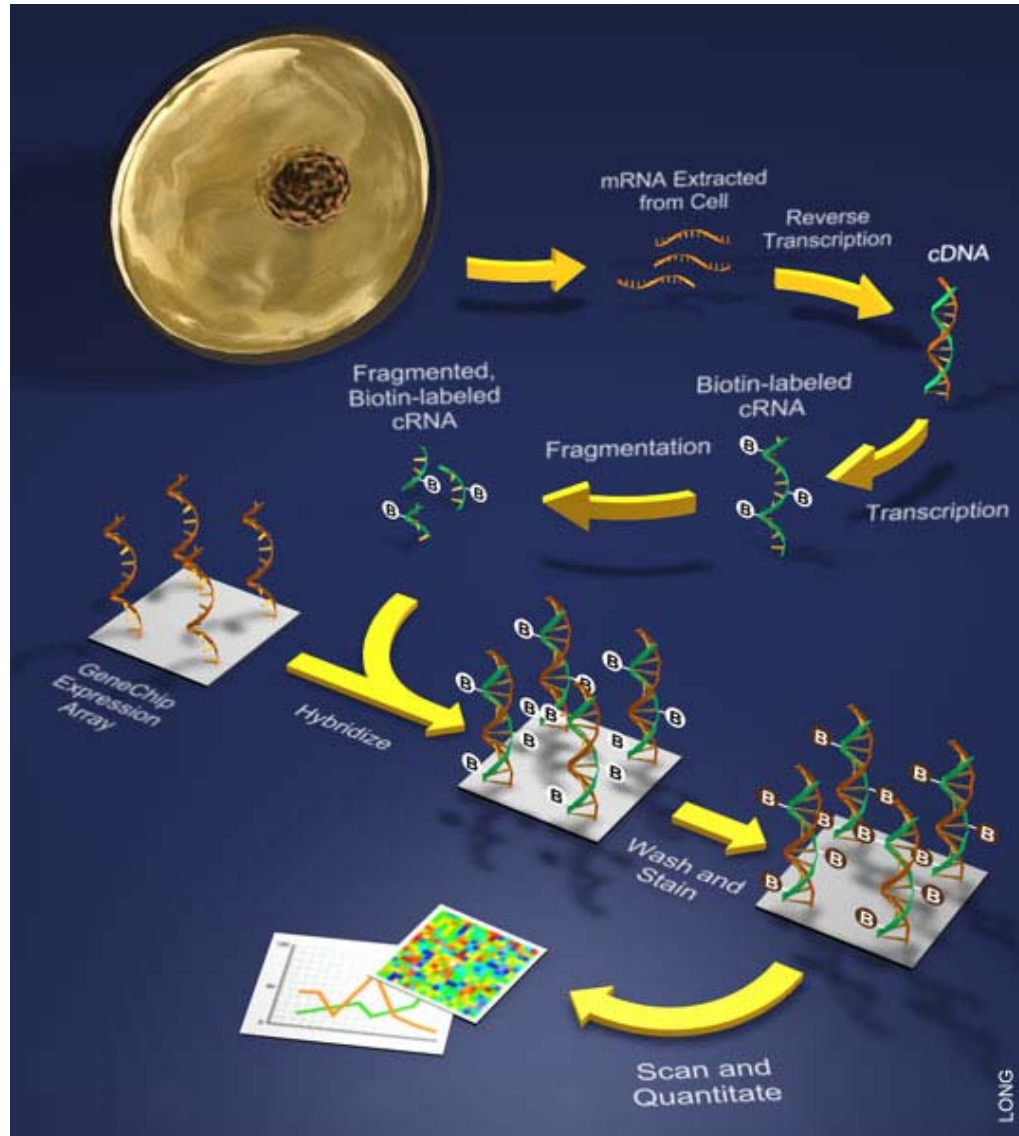


Cell biology & proteomics data...

protein-protein
interactions



Transcriptomics: DNA microarray technology



Phenotype data: human diseases

OMIM - ACHOO SYNDROME - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address <http://www.ncbi.nlm.nih.gov/entrez/dispmim.cgi?id=100820> Go Links

NCBI

MIM %100820
Text
References
Contributors
Creation Date
Edit History

Clinical Synopsis

OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search OMIM for Go Clear

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

All: 1 OMIM dbSNP: 0 OMIM UniSTS: 0

%100820
ACHOO SYNDROME

Links

Alternative titles; symbols

AUTOSOMAL DOMINANT COMPELLING HELIOOPHTHALMIC OUTBURST SYNDROME
PHOTIC SNEEZE REFLEX
SNEEZING FROM LIGHT EXPOSURE
PEROUTKA SNEEZE

TEXT

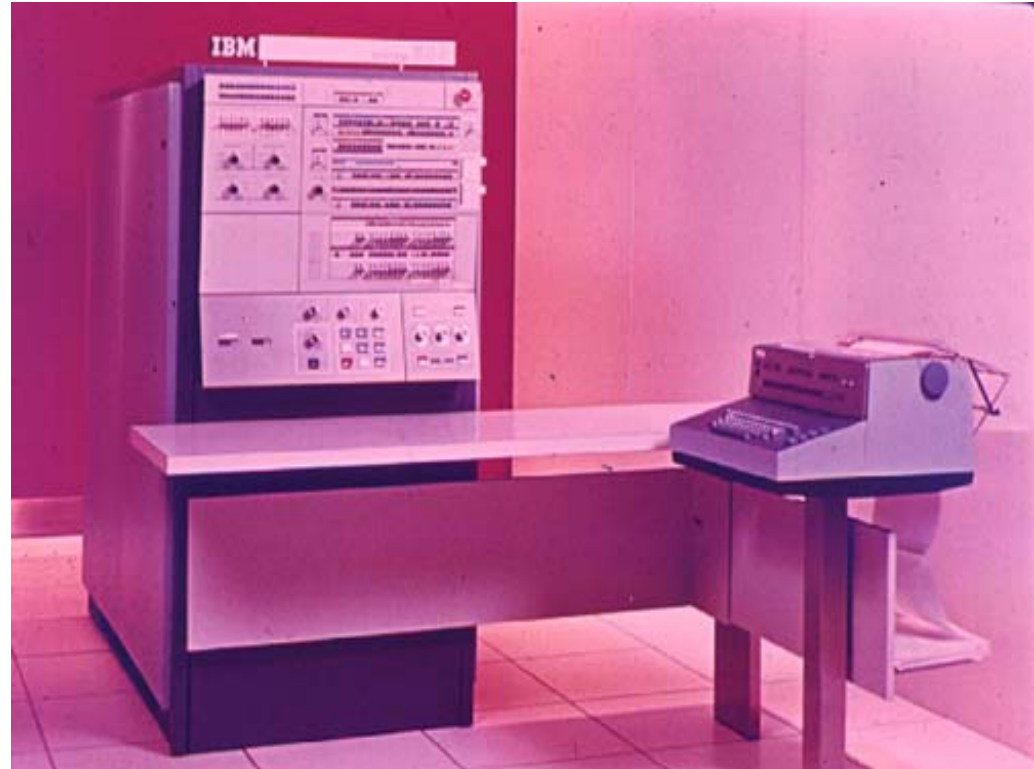
[Collie et al. \(1978\)](#) described a 'disorder' characterized by nearly uncontrollable paroxysms of sneezing provoked in a reflex fashion by the sudden exposure of a dark-adapted subject to intensely bright light, usually sunlight. The number of successive sneezes was usually 2 or 3, but could be as many as 43. The 4 authors were the probands of the 4 families they reported. Several instances of male-to-male transmission were noted. Sneezing in response to bright light was said by [Peroutka and Peroutka \(1984\)](#) to be a common yet poorly understood phenomenon. Photoc sneeze reflex was suggested as the appropriate designation by [Everett \(1964\)](#), who found it in 23% of Johns Hopkins medical students. In a poll of 25 neurologists at Johns Hopkins, [Peroutka and Peroutka \(1984\)](#) found the phenomenon in 9, but only 2 of the respondents knew that such a specific reflex exists. The Peroutkas (father and daughter) reported the reflex in 3 generations of their family: grandfather, the father (the proband), his brother and his daughter. The index subject (S.J.P.) invariably sneezes twice when he moves from indoors into bright sunlight. [Lewkonia \(1969\)](#) described sneezing as a complication of slit lamp examination.

Internet

- Homology / Alignment
- Simple pattern (“word”) recognition
- Statistical methods
 - Weight matrices: calculate amino acid *probabilities*
 - *Other examples:* Regression, variance analysis, clustering
- Machine learning
 - Like statistical methods, but parameters are estimated by iterative *training* rather than direct calculation
 - *Examples:* Neural Networks (**NN**), Hidden Markov Models (**HMM**), Support Vector Machines (**SVM**)
- Combinations

The computer

- *Everything* can be reduced to bits (0 or 1)



- A byte = 8 bits

0 1 0 0 0 0 0 1

Can be interpreted as

- The number 65
- The letter "A"
- Part of a machine code instruction
- Part of a colour specification
- Part of a sound encoding
- ...

A text file is a file
where every byte is
interpreted as a
character

Examples

Plain text	.txt
Program settings	.ini
C source code	.c
Python script	.py
T _E X source	.tex
Web page source	.html
Sequences	.fasta

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	@	96	60	`
1	01	Start of heading	33	21	!	65	41	A	97	61	a
2	02	Start of text	34	22	"	66	42	B	98	62	b
3	03	End of text	35	23	#	67	43	C	99	63	c
4	04	End of transmit	36	24	\$	68	44	D	100	64	d
5	05	Enquiry	37	25	%	69	45	E	101	65	e
6	06	Acknowledge	38	26	&	70	46	F	102	66	f
7	07	Audible bell	39	27	'	71	47	G	103	67	g
8	08	Backspace	40	28	(72	48	H	104	68	h
9	09	Horizontal tab	41	29)	73	49	I	105	69	i
10	0A	Line feed	42	2A	*	74	4A	J	106	6A	j
11	0B	Vertical tab	43	2B	+	75	4B	K	107	6B	k
12	0C	Form feed	44	2C	,	76	4C	L	108	6C	l
13	0D	Carriage return	45	2D	-	77	4D	M	109	6D	m
14	0E	Shift out	46	2E	.	78	4E	N	110	6E	n
15	0F	Shift in	47	2F	/	79	4F	O	111	6F	o
16	10	Data link escape	48	30	0	80	50	P	112	70	p
17	11	Device control 1	49	31	1	81	51	Q	113	71	q
18	12	Device control 2	50	32	2	82	52	R	114	72	r
19	13	Device control 3	51	33	3	83	53	S	115	73	s
20	14	Device control 4	52	34	4	84	54	T	116	74	t
21	15	Neg. acknowledge	53	35	5	85	55	U	117	75	u
22	16	Synchronous idle	54	36	6	86	56	V	118	76	v
23	17	End trans. block	55	37	7	87	57	W	119	77	w
24	18	Cancel	56	38	8	88	58	X	120	78	x
25	19	End of medium	57	39	9	89	59	Y	121	79	y
26	1A	Substitution	58	3A	:	90	5A	Z	122	7A	z
27	1B	Escape	59	3B	;	91	5B	[123	7B	{
28	1C	File separator	60	3C	<	92	5C	\	124	7C	
29	1D	Group separator	61	3D	=	93	5D]	125	7D	}
30	1E	Record separator	62	3E	>	94	5E	^	126	7E	~
31	1F	Unit separator	63	3F	?	95	5F	_	127	7F	□

The ASCII table

Extended character sets

! " # \$ % & ' () * + , - . /
 0 1 2 3 4 5 6 7 8 9 : ; < = > ?
 @ A B C D E F G H I J K L M N O
 P Q R S T U V W X Y Z [\] ^ _
 ` a b c d e f g h i j k l m n o
 p q r s t u v w x y z { | } ~ .
 € . , f „ … † ‡ ^ % Š < € . Ž .
 . ' ' " " • — — ~ ™ Š > œ . ž Ÿ
 ; ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ ® ¯
 ° ± ² ³ ´ µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿
 À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï
 Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß
 à á â ã ä å æ ç è é ê ë ì í î ï
 ð ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ

Windows-1252, sometimes called incorrectly "ANSI". Blue dots indicate unused or control characters

The are *many* ways to interpret characters with values above 127. Here, you see two of them.

"Mac OS Roman" Encoding:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
032		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
048	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
064	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
080	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
096	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
112	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
128	Ä	Å	Ç	É	Ñ	Ö	Ü	á	à	â	ã	ä	å	ç	é	è
144	ê	ë	í	ì	ï	î	ñ	ó	ò	ô	õ	ö	ù	û	ü	
160	†	°	¢	£	§	•	¶	ß	®	©	™	ˆ	˜	≠	Æ	Ø
176	∞	±	≤	≥	¥	μ	∂	Σ	Π	π	∫	ª	º	Ω	æ	ø
192	¿	¡	¬	√	ƒ	≈	Δ	«	»	...		À	Á	Ï	Œ	œ
208	–	—	“	”	‘	’	÷	◊	ÿ	ÿ	/	€	<	>	fi	fl
224	‡	·	,	„	‰	Â	Ê	Á	Ë	È	Í	Î	Ï	Ì	Ó	Ô
240	Apple	Ò	Ú	Û	Ü	ı	ˆ	-	-	˘	-	˙	˚	˛	˜	˘

Text files—line endings

- UNIX standard (including Mac OS X):
 - 10 — LF ("Line feed" char).
- Old Mac (System 9 and before):
 - 13 — CR ("Carriage Return" char).
- DOS/Windows:
 - 13, 10 — both CR and LF.

A good text editor can handle all three systems.

Notepad for Windows *cannot!*

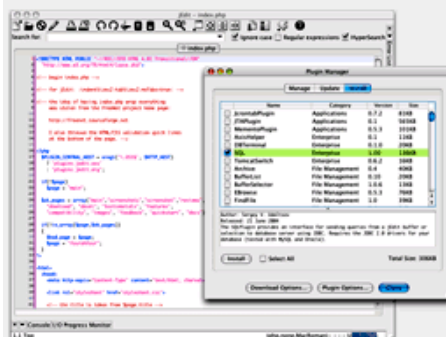
Last Site Update: 19 November 2011 | Latest Version: 4.5pre1 | Stable Version: 4.4.2

jEdit

Programmer's Text Editor



Download



jEdit is a mature programmer's text editor with hundreds (counting the time developing plugins) of person-years of development behind it. To [download](#), install, and set up jEdit as quickly and painlessly as possible, go to the [Quick Start](#) page.

While jEdit beats many expensive development tools for features and ease of use, it is released as free software with full source code, provided under the terms of the [GPL 2.0](#).

The jEdit core, together with a large collection of [plugins](#) is maintained by a [world-wide developer team](#).

Some of jEdit's features include:

- Written in Java, so it runs on Mac OS X, OS/2, Unix, VMS and Windows.
- Built-in macro language; extensible plugin architecture. Hundreds of macros and plugins available.
- Plugins can be downloaded and installed from within jEdit using the "plugin manager" feature.
- Auto indent, and syntax highlighting for more than 200 languages.
- Supports a large number of character encodings including UTF8 and Unicode.
- Folding for selectively hiding regions of text.
- Word wrap.
- Highly configurable and customizable.
- Every other feature, both basic and advanced, you would expect to find in a text editor. See the [Features](#) page for a full list.



About

- Main Site
- Features
- Compatibility
- Screenshots
- Icons and Images
- Reviews
- Download
- Plugins

Community

- jEdit Community

Help

- Quick Start Guide
- Online Documentation
- Feedback and Support

Development Links

- Development
- SourceForge Project

sourceforge

JProfiler

Donate

